Predicting
Observation Impact on Forecast Skill
with Machine Learning

Francois Vandenberghe & Dick Dee

Joint Center for Satellite Data Assimilation (JCSDA)

Boulder, CO

vandenb@ucar.edu

Assimilation cost function:

$$J = \left(x - x^b\right)^T B^{-1}\left(x - x^b\right) + (y^o - Hx)^T R^{-1}(y^o - Hx)$$

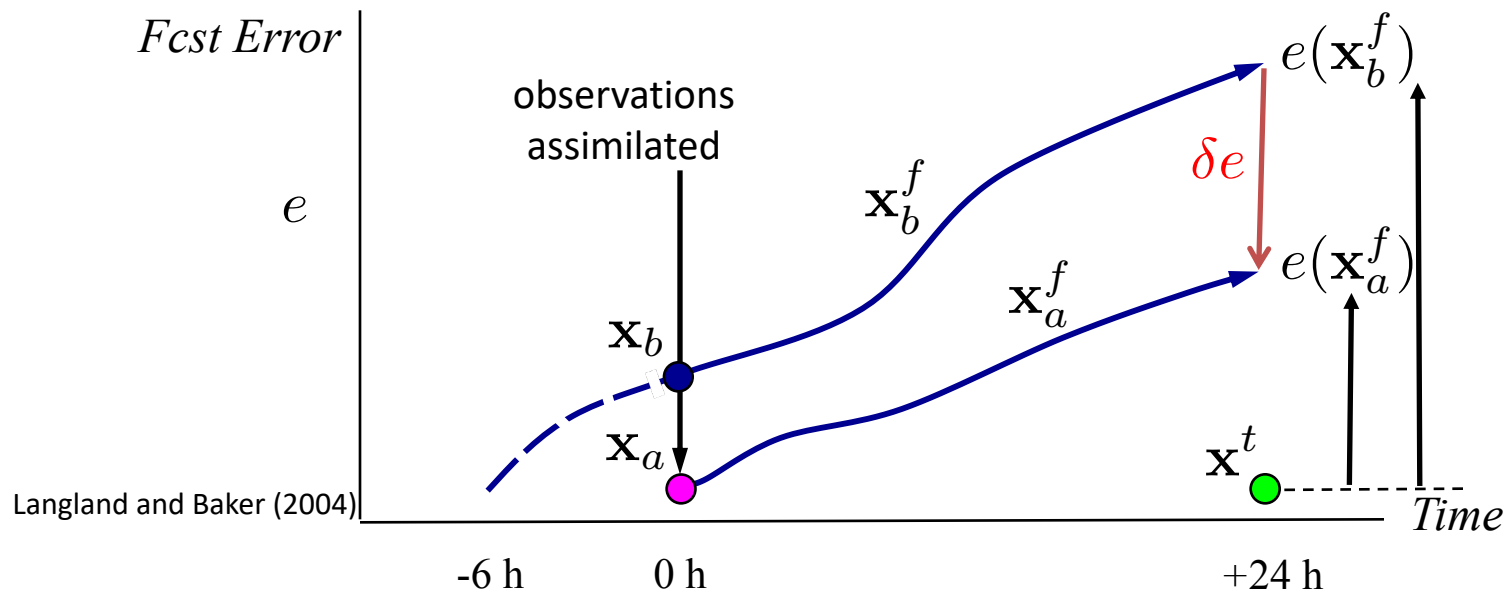**Analysis**

$$x^a = x^b + \underbrace{BH^T(HBH^T + R)^{-1}}_{K}\underbrace{\left(y^o - Hx^b\right)}_{d}$$

Analysis uncertainty

$$A = (I - KH)B$$

Langland and Baker (2004)

Adjoint-derived (single outer-loop) observation impact

$$\delta e \approx d^T K^T [M_b^T e(x_b^f) + M_a^T e(x_a^f)]$$
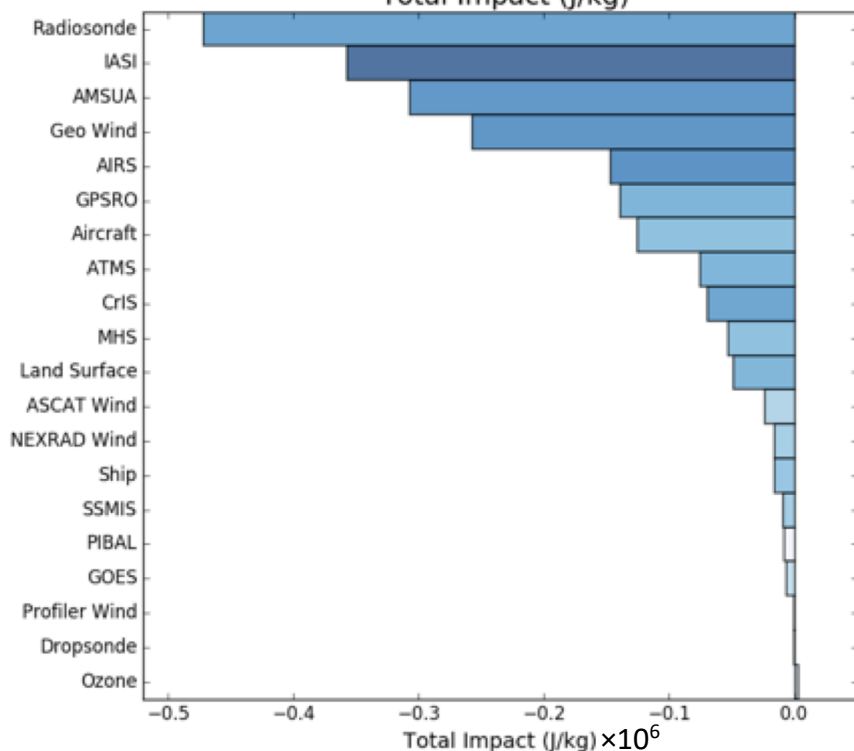
Ensemble-derived observation impact

$$\delta e \approx d^T R^{-1} L(HX_a^0) X_a^{fT} [e(x_b^f) + e(x_a^f)]$$
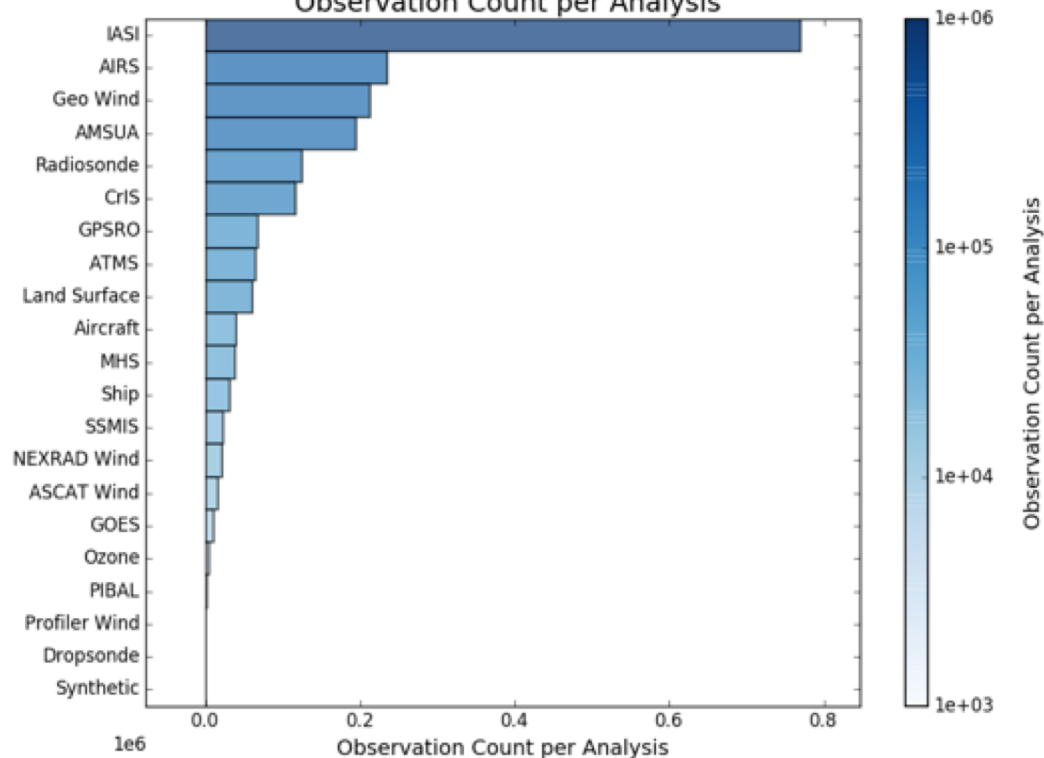
Forecast Sensitivity to Observations Impact (FSOI)

Total Impact

EMC 24h Observation Impact Summary
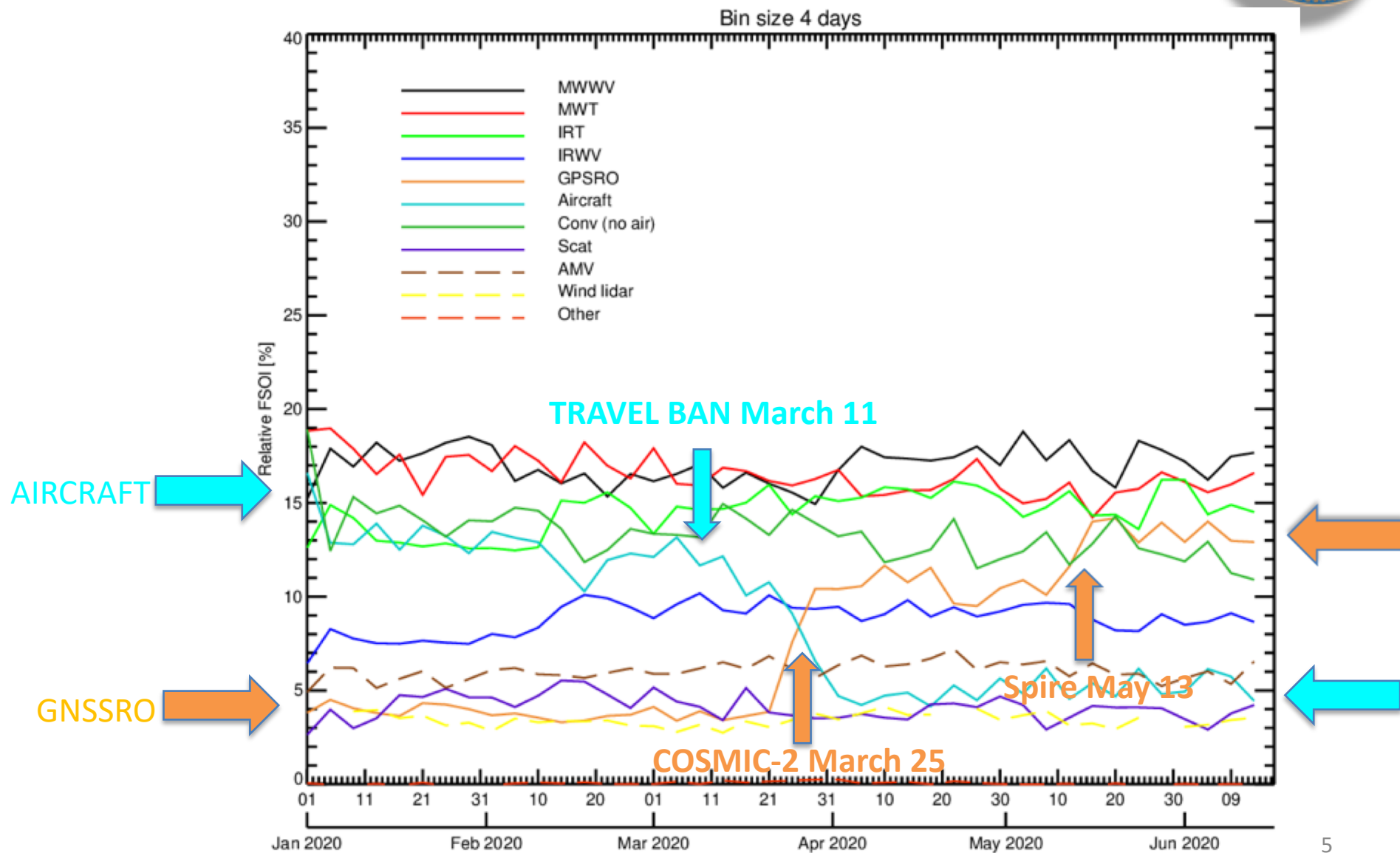Global, 00Z 07Apr-06May,2018
Total Impact (J/kg)

Observations count

EMC 24h Observation Impact Summary
Global, 00Z April 2018
Observation Count per Analysis

Fractional Impact

Seek for the radiance bias correction that maximizes FSOI:

- **Use Machine Learning to compute bias correction coefficients (MLBC)**

- **Do not limit the number of predictors. (big data)**

$$I = \left(y^o - Hx^b\right)^T \frac{\partial F}{\partial y}$$

impact   innovation   sensitivity

Can Forecast Sensitivity to Observations be predicted?

Machine Learning

Software: **TenserFlow on Amazon**

Data**:**   **Dec 2014, Jan 2015 & Feb 2015**

      Focus initially on AMSU

Training**:**   **Dec 1 – Feb 14**

Prediction**:**   **Feb 14 - 28**

Predictors: **all FV3 first guess variables:**

2D: Topography geopotential, Surface temperature, Fraction-of-land, Fraction-of-land-ice, Fraction-of-lake, Fraction-of-ocean, Fraction-of-ocean-ice, Surface Pressure. **3D**: Pressure Thickness, Zonal Wind, Meridional wind, Virtual Temperature, Specific Humidity, ozone, Mass Fraction Cloud Ice Water, Mass Fraction Cloud Liquid Water.
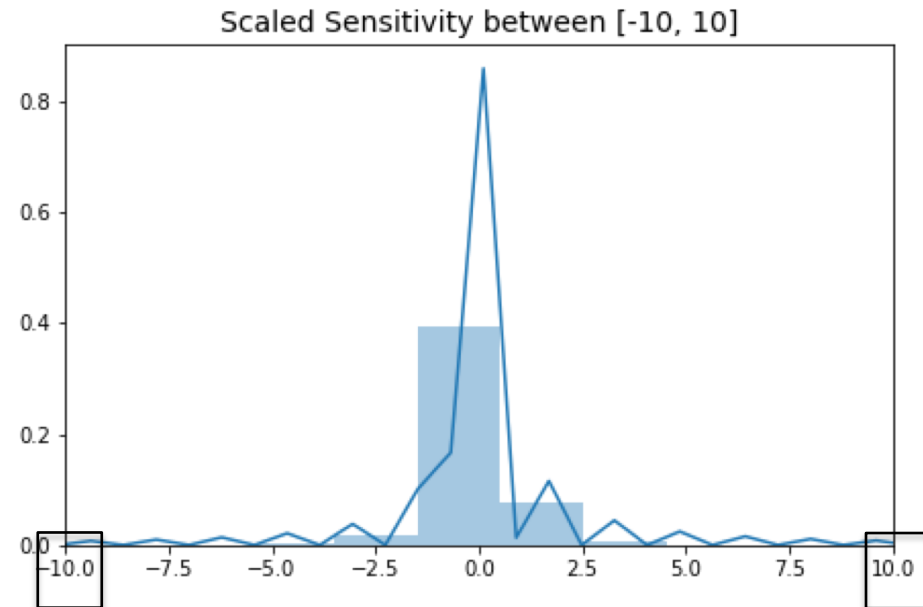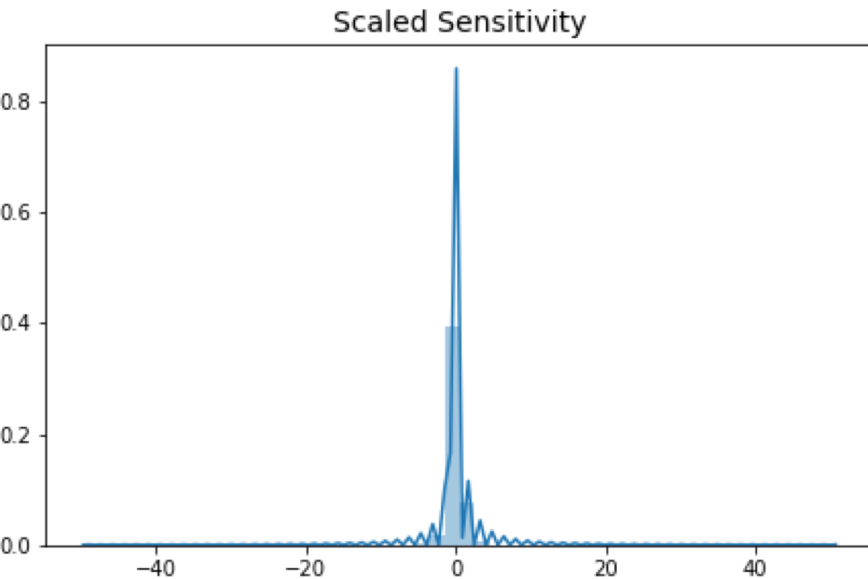
Data Size

	Number of files	Avg file size	Total files size
Observation	360	48.9 MB	17.2 GB
Model background	360	464.2 MB	163.2 GB

Data Analysis

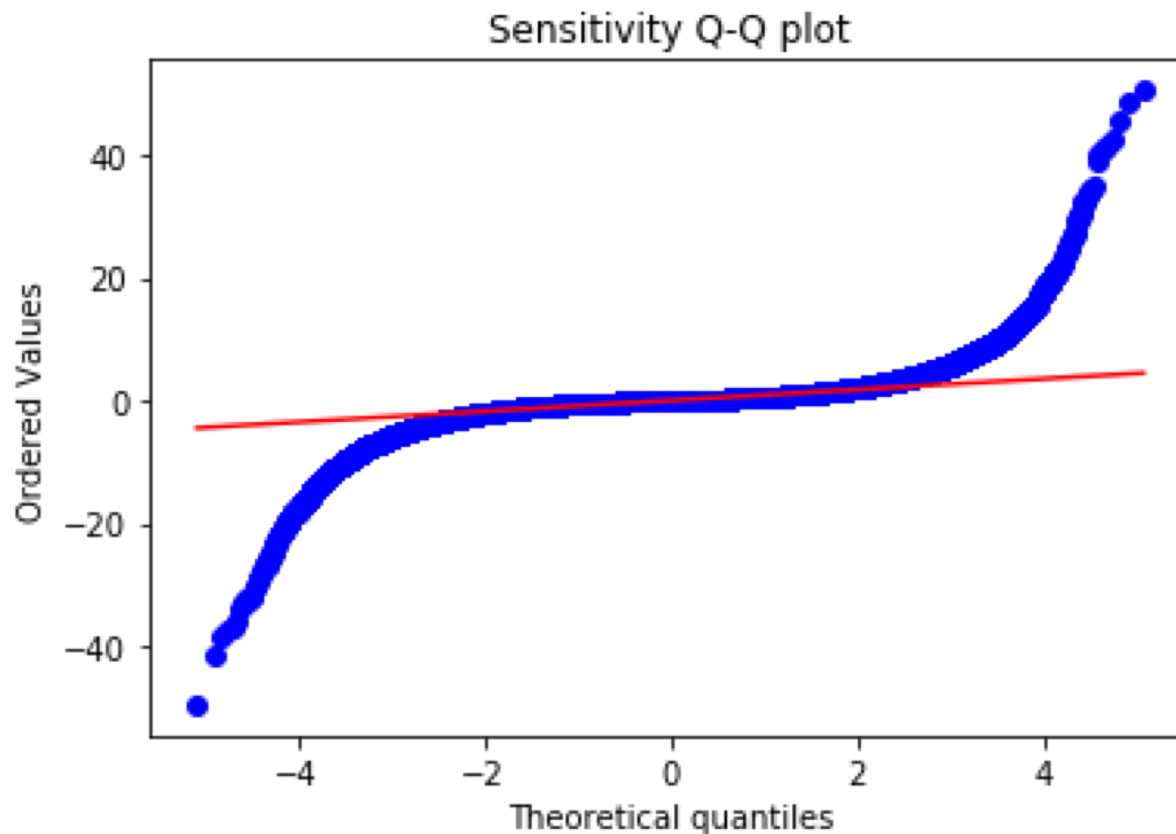## GMAO Dec-Jan-Feb 2015 scaled sensitivity AMSU N18 channel 7

p-value = 0.0, sample does not look Gaussian

GMAO Dec-Jan-Feb 2015 scaled sensitivity AMSU N18 channel 7

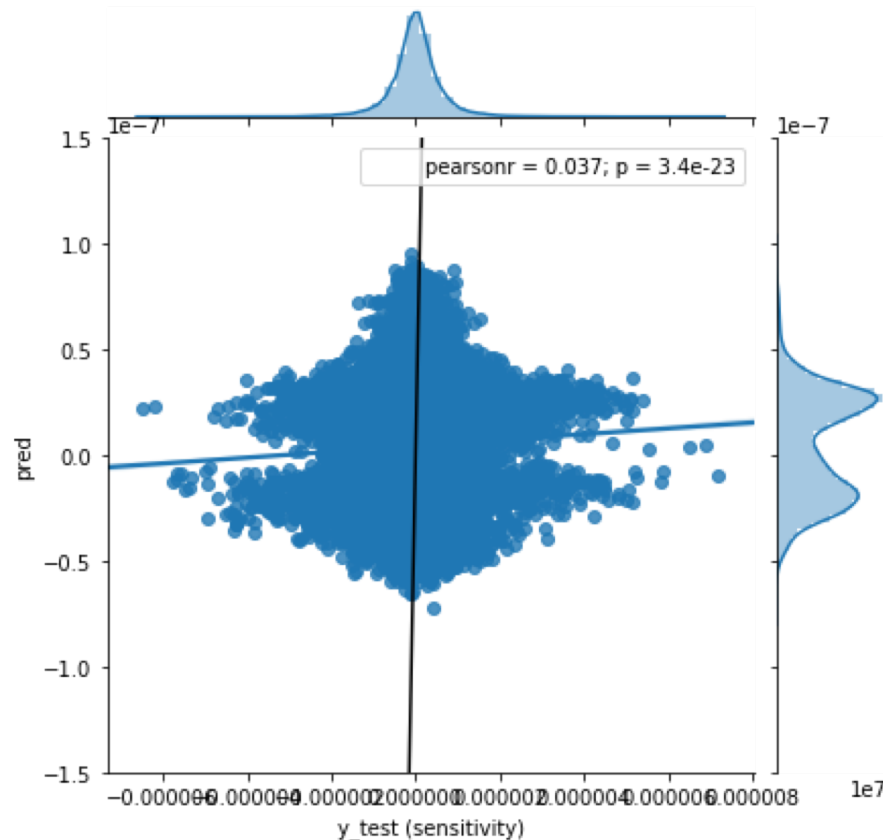Sensitivity Q-Q plot
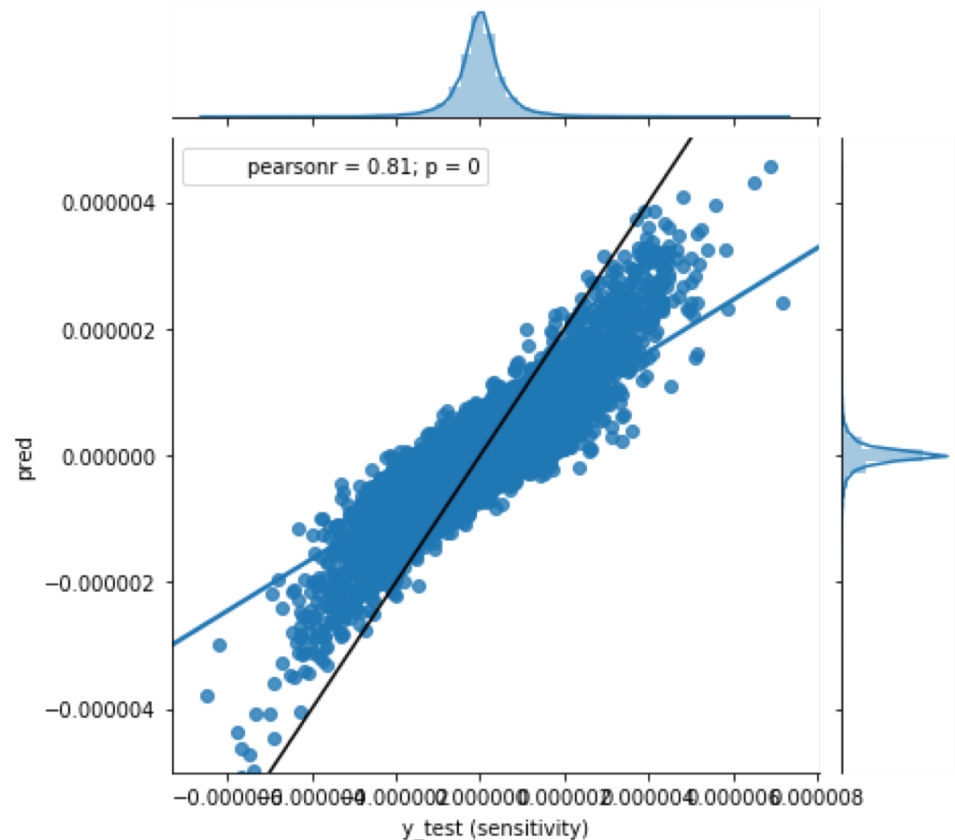
Kurtosis measure: 54.68

Machine Learning

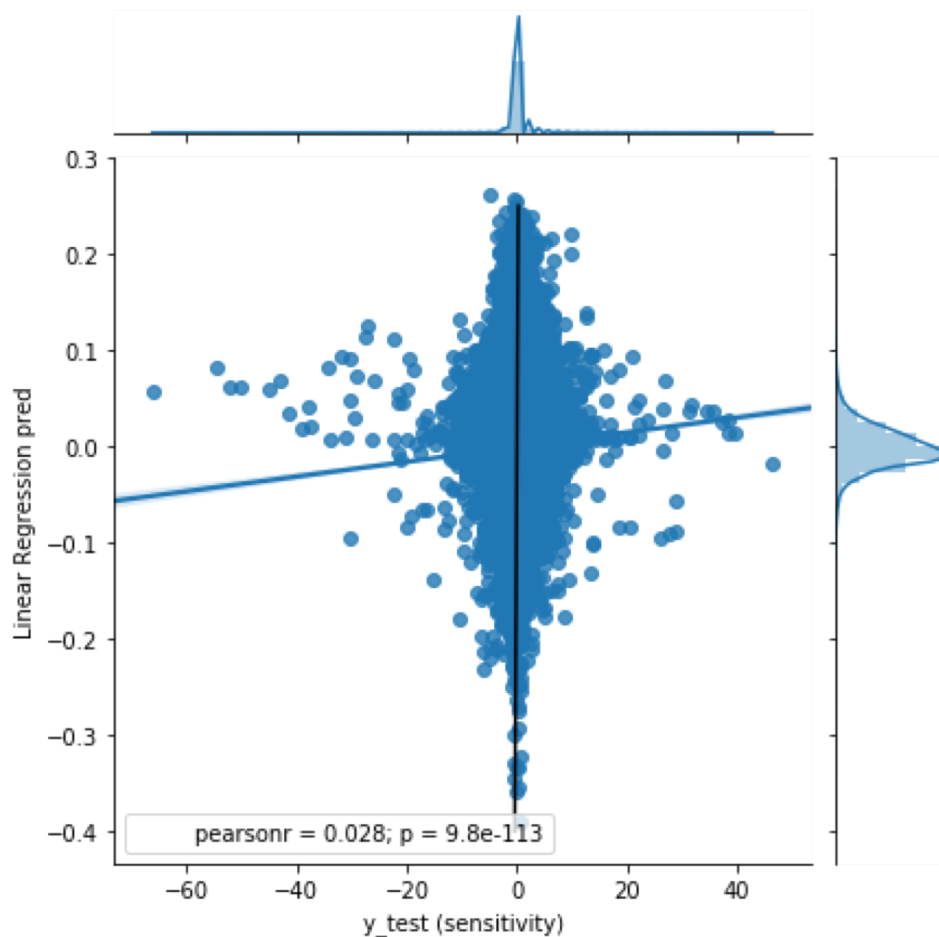**Linear regression**   **Gradient Boosting**

GMAO Forecast Sensitivity to AMSUA Channel 14:
Test data (x-axis) and predictions (y-axis)
00z analyses for December 2014 (~3.5x10^5 obs)

Linear Regression

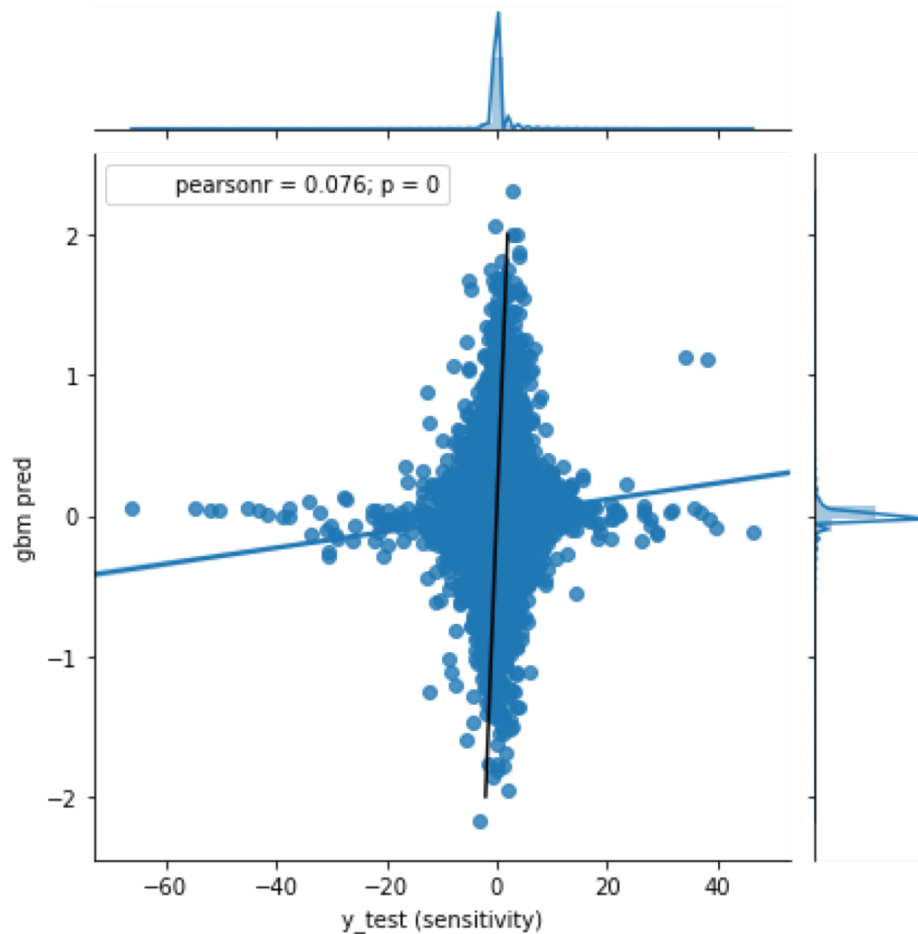**Linear Regression**

Test data (x-axis) and predictions (y-axis)
for GMAO Dec-Jan-Feb 2015 sensitivity AMSU N18 channel 7

Gradient Boosting

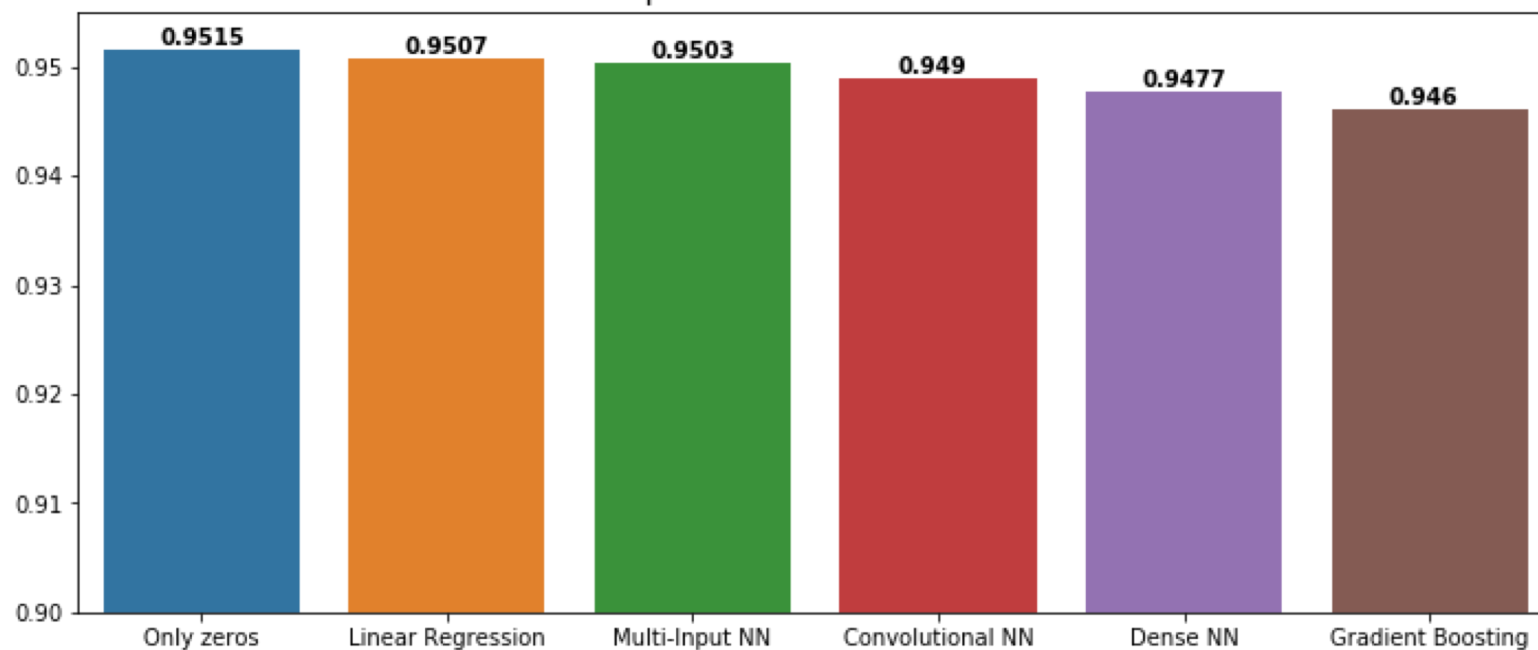**Gradient Boosting**

Test data (x-axis) and predictions (y-axis)
for GMAO Dec-Jan-Feb 2015 sensitivity AMSU N18 channel 7

Results

Mean Squared Errors AMSUA N18 CH7

Summary

- **Forecast to Observation Sensitivity is not gaussian.**

- **Sophisticated ML techniques lead to better results than simple regression**

- **Very limited predictability.**